Fuel 97 (2012) 710–717

Contents lists available at SciVerse ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system

Julio Cesar L. Alves [a], Claudete B. Henriques [b], Ronei J. Poppi [a,*]

[a] Institute of Chemistry, University of Campinas – UNICAMP, P.O. Box 6154, 13083-970 Campinas, SP, Brazil
[b] Planalto Refinery-REPLAN – Petrobras, Paulinia, SP, Brazil

ARTICLE INFO

Article history:
Received 29 November 2011
Received in revised form 7 March 2012
Accepted 11 March 2012
Available online 27 March 2012

Keywords:
Diesel
Near infrared spectroscopy
Support vector regression
In-line blending

ABSTRACT

This work demonstrates the application of support vector regression (SVR) applied to near infrared spectroscopy (NIR) data to solve regression problems associated to determination of quality parameters of diesel oil for an in-line blending optimizer system in a petroleum refinery. The determination of flash point and cetane number was performed using SVR and the results were compared with those obtained by using the PLS algorithm. A parametric optimization using a genetic algorithm was carried out for choice of the parameters in the SVR regression models. The best models using SVR presented a RBF kernel and spectra preprocessed with baseline correction and mean centered data. The obtained values of RMSEP with the SVR models are 1.98 °C and 0.453 for flash point and cetane number, respectively. The SVR provided significantly better results when compared with PLS and in agreement with the specification of the ASTM reference method for both quality parameter determinations.

1. Introduction

In several refineries, semi-finished or finished products (in particular commercial fuels, such as diesel oil) are not directly extracted from parts of crude oils, but are produced by blending several components. As shown in Fig. 1, an in-line blending system for diesel oil is the process of combining a number of feedstocks, produced by other refinery process units, together with small amounts of additives, to make a mixture meeting certain quality specifications [1,2].

In a blending process, several components are pumped to a blender (also referred to as static mixer) from intermediate storage tanks or pipes. After passing by the blender, the product is stored in a final product tank, routed to another refining unit or shipped off the refinery. Due to the fact that is the final stage in a refinery process, the optimization of this process is of vital importance. Regardless of how efficient the upstream process units may be, this can be invalid if poorly optimized blending produces a substandard fuel. In many respects it is the most important process to optimize and it can also bring the maximum benefits in terms of quality of product and payback [1,2].

Automatic analyzers based on near infrared spectroscopy (NIR) [3,4] for in-line blending diesel optimization allows for rapid multi-stream and multi-property quality parameter determinations of diesel blending components and final product streams [5–9]. The calibration methodologies used and the transferability of calibrations between laboratory analyzers and process blending analyzers allow for rapid project startup, and minimize the amount of site-specific calibration work that is needed. For these reasons, the cost of monitoring can be significantly reduced, compared with conventional final product blending analysis methods [8–11].

The rapidity of the methods and the quality of results obtained by the calibration models allows their utilization for on-line parameter determination, providing sensitive productivity improvement when used in process control, where measurements need to be fast and accurate to allow a central system response in a case of disturbances or set point tendencies that differ from desired values [3,4,8–11].

Also, combined with the flexibility in data acquisition made possible by near infrared spectroscopy, regression model development with the capacity to properly adjust a possible nonlinear relationship and with a high generalization performance is necessary [11–15]. This is due to the fact that NIR spectroscopy data of petroleum derivative samples may present nonlinear correlations [16,17] with certain quality parameters, depending on the analytical range used. Due to the variability of crude oils processed in the refinery and the different streams that make up the in-line blending of diesel oil, this analytical range should be as comprehensive as possible. Moreover, even taking these precautions, it is possible to need the prediction of a sample that is outside the analytical range included in the model, and in this case its generalization performance must be appropriate to prediction with minor error [11–15].

Diesel oil has several limiting properties, which prevent certain streams to be indiscriminately added in the blend process, since

* Corresponding author.
E-mail address: ronei@iqm.unicamp.br (R.J. Poppi).

0016-2361
http://dx.doi.org/10.1016/j.fuel.2012.03.016

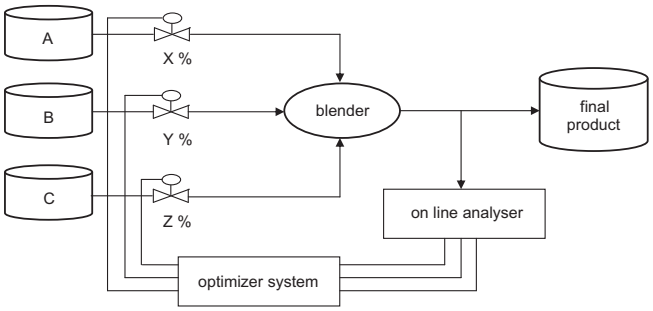

Fig. 1. Typical in-line blending process of diesel oil.

the additions may contribute in an inadequate way to the specified set point. This is the case of flash point and cetane number, which limits, for example, the addition of very light fractions such as heavy naphtha.

Flash point measures the tendency of the specimen to form a flammable mixture with air under controlled laboratory conditions and it is used in shipping and safety regulations to define flammable and combustible materials. Flash point can indicate the possible presence of highly volatile and flammable materials in a relatively nonvolatile or nonflammable material. For example, an abnormally low flash point on a sample of diesel oil can indicate gasoline contamination. In a blending process of diesel oil the flash point can indicate an excessive amount of light fractions.

The cetane number provides a measure of the ignition characteristics of diesel fuel oil in compression ignition engines. This test method is used by engine manufacturers, petroleum refiners and marketers, and in commerce as a primary specification measurement related to matching of fuels and engines.

Thus, this study demonstrates the use of the support vector regression algorithm applied to NIR spectroscopy data to obtain calibration models that can be applied for determination of diesel quality parameters for an in-line blending optimizer system in a petroleum refinery.

1.1. Support vector regression

Partial least squares (PLS) [18] is currently the most popular algorithm for multivariate calibration development, due to its ease of both implementation and interpretation of results. It is a linear multivariate calibration method that is able to model soft nonlinearities by appropriate choices of the number of latent variables [18]. Recently, applications of support vector machine algorithms have demonstrated a sensitive improvement in results compared to that obtained with PLS [19,20], especially for data with high nonlinear relationships or complexities.

Table 1
Results for PLS and SVM models.

	Flash point			Cetane number		
	RMSEC (°C)	RMSEP (°C)	R^2	RMSEC	RMSEP	R^2
PLS	4.21	3.77	0.698	0.745	0.556	0.894
SVR	1.99	1.98	0.936	0.765	0.453	0.895

Artificial neural networks (ANNs) has also been used for nonlinear multivariate calibration, however the final results depend on the initial parameters, sometimes is necessary to repeat the network training, the final solution is non-unique and it has the tendency to overfit. SVR has the advantage in relation to ANN in produce a global model that is capable of efficiently dealing with non-linear relationships [8].

The support vector machines [21,22] are learning machines that are based on statistical learning theory, trained through a supervised learning algorithm. Support vector machines for regression, such as support vector regression (SVR) [23,24] are based on the estimation of the function:

$$f(\mathbf{x}) = (\mathbf{w}\phi(\mathbf{x}) + b) \tag{1}$$

where the input vectors **x** are mapped into a high-dimensional feature space **Z** through some nonlinear mapping, Φ: xi → zi; chosen a priori.

In the case of SVR, the regression parameters are calculated by minimizing:

$$\frac{1}{2}\|\mathbf{w}\|^2 + CR_{emp} \tag{2}$$

where R_{emp} is the empirical risk (or training error) and C is a regularized parameter which determines the relationship in minimizing the training error and the model complexity term $\|\mathbf{w}\|^2$.

Using the so called ε-insensitive loss function chosen a priori:

$$|y - f(x_i)|_\varepsilon = \max\{0, |y_i - f(x_i) - \varepsilon|\} \tag{3}$$

The empirical risk can be calculated as:

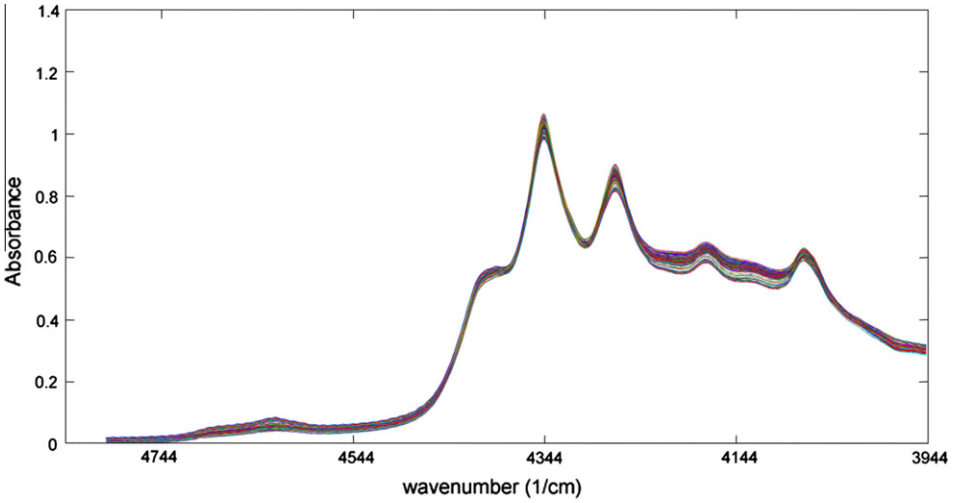

Fig. 2. Spectra of 451 diesel oil samples for flash point calibration model.

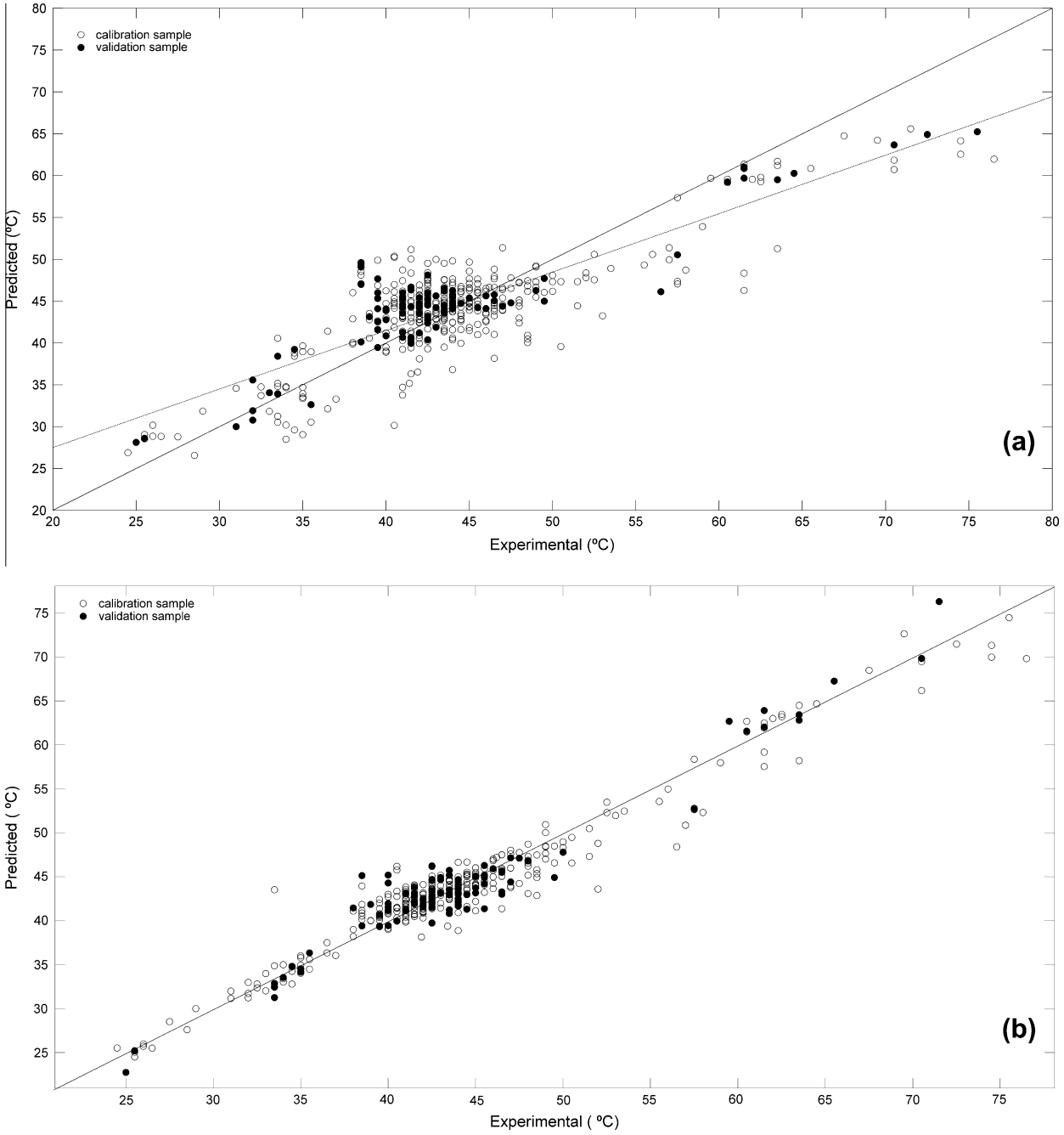

Fig. 3. Flash point calibration model. Calibration (○) and validation (●) for PLS (a) and SVR (b).

$$R_{emp} = \frac{1}{n}\sum_{i=1}^{n}|y_i - f(x_i)|, \varepsilon \tag{4}$$

Examining Eq. (3) it is possible to observe that data points that lie inside a tube with radius ε do not contribute to the solution. On the other hand, the points lying outside the ε tube are named support vectors, because these establish the fundaments of the estimated regression function.

The slack variables ξ and ξ^* are introduced for the situation that the point exceeds the ε-sensitive zone. Thus, the ε-SVR is equivalent to solving the following constrained optimization problem:

$$\text{minimize}: \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\frac{1}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{5}$$

subject to the following constraints:

$$f(x_i) - y_i \leqslant \varepsilon + \xi_i, \tag{6}$$

$$y_i - f(x_i) \leqslant \varepsilon + \xi_i^* \tag{7}$$

$$\varepsilon, \xi_i, \xi_i^* \geqslant 0 \tag{8}$$

Although the parameter ε controls the sparseness of the solution, it does this in an indirect way. Since we do not know the information about the accuracy of the y-values, it can be difficult to find a reasonable value of ε a priori. Instead, it is necessary to specify the degree of sparseness and the algorithm must compute ε from the data. This is the idea of v-SVR [25], a modification of the original ε-SVR, where a parameter v controls the number of support vectors and the number of points that come to lie outside of the ε-insensitive tube. It promotes the highest generalization of

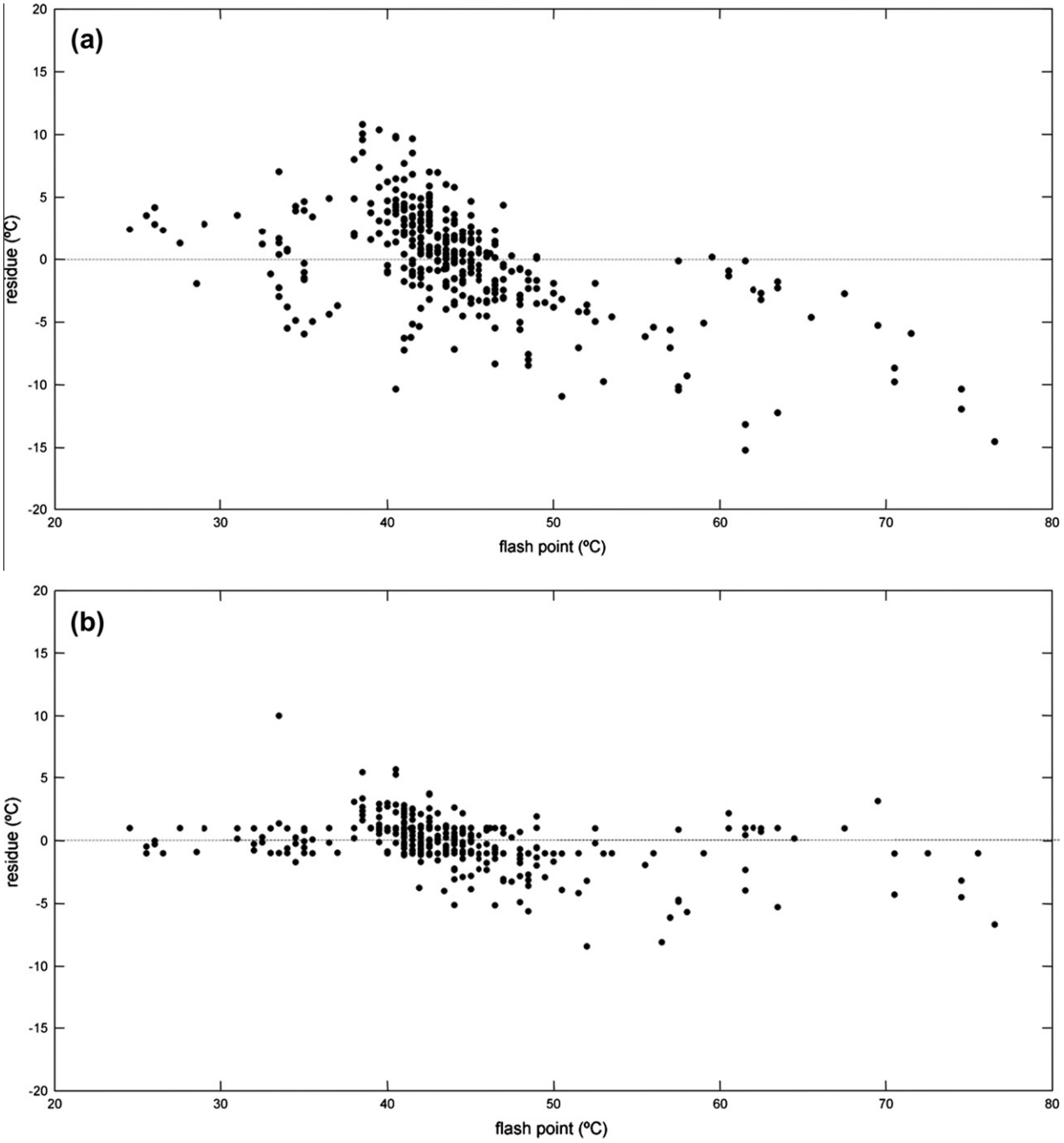

Fig. 4. Residual plots of flash point models for PLS (a) and SVR (b).

the model and the parameter v is found in accordance with the noise in the y-values and can be chosen automatically by an optimization algorithm. In this particular v-SVR application a genetic algorithm was used to select the SVR parameters (v and C), but type of kernel function and its parameters were kept constant.

1.2. Parametric optimization with genetic algorithm – GA

The concept of genetic algorithm (GA) was developed by Holland and his colleagues in the 1960s and 1970s [26,27]. GA are inspired by the evolutionist theory explaining the origin of species and is a stochastic search technique which can be used to find the global optimal solution in a complex multidimensional search space.

In GA terminology, a solution vector is called an individual or a chromosome, which are made of discrete units called genes and each gene represents the actual parameters to be optimized. GA work on the encoding of a problem, not on the problem itself. Conventionally, the chromosomes in a GA are binary coded which in fact lead to integer valued solutions.

GA operate with a collection of chromosomes, called a population. The population is normally randomly initialized. As the search evolves, the population includes fitter and fitter solutions, and eventually it converges, meaning that it is dominated by a single solution. The solutions from one population are used to generate the next population. In order to create a new population GA uses two operators: crossover and mutation. Genetic operators are used to generate the new solutions (children population or offspring)

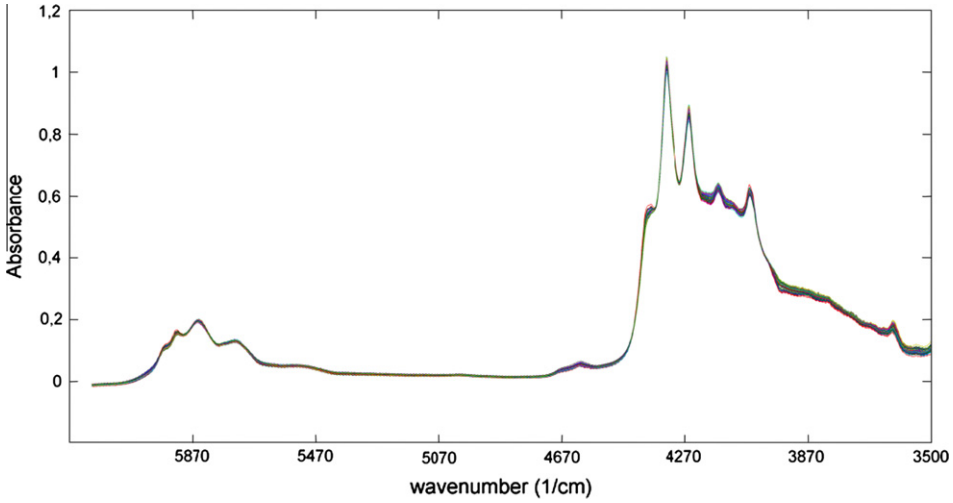

Fig. 5. Spectra of 114 diesel oil samples for cetane number calibration model.

from the current set of solutions (parent population). Selection reflects the principle of survival of the fittest and is the driving mechanism of keeping and deleting some solutions from the parent population to generate an offspring with the same number of chromosomes. During this selection process, the solutions are selected according to their values of objective function. The GA will repeat this process until a termination condition is satisfied [23,28].

The procedure of a generic GA is given as follows:

1. Choose a randomly generated population.
2. Calculate the fitness of each chromosome in the population.
3. Create the offspring by genetic operators.
4. Check the termination condition. If the new population does not satisfy the termination condition, repeat steps 2 up to 4 for the generated offspring as a new starting population.

2. Method

For calibration models the procedures described in ASTM D56 and ASTM D613 for flash point and cetane number determination, respectively, were performed and the NIR spectra was obtained.

In the ASTM D56 the specimen is placed in the cup of the tester and, with the lid closed, heated at a slow constant rate. An ignition source is directed into the cup at regular intervals. The flash point is taken as the lowest temperature at which application of the ignition source causes the vapor above the specimen to ignite. For flash point determination (tag closed cup method) an FP56 5G2 ISL automatic analyzer was used.

In the ASTM D613 the cetane number of a diesel fuel oil is determined by comparing its combustion characteristics in a test engine with those for blends of reference fuels of known cetane number under standard operating conditions. This is accomplished using the bracketing handwheel procedure which varies the compression ratio (handwheel reading) for the sample and each of two bracketing reference fuels to obtain a specific ignition delay permitting interpolation of cetane number in terms of handwheel reading. The cetane number determination was carried in a Waukesha standard engine.

The NIR spectra was obtained with an Bomen MID/NIR spectrometer with Glowbar source, DTGS detector, using a transmittance sample cell of CaF_2 with 0.5 mm optical path. Each spectrum was obtained as the average of 32 scans, with resolution of 2 cm^{-1}.

Models were developed using 451 samples (350 calibration and 101 validation samples) for flash point model and 114 samples for cetane number determination (77 calibration and 37 validation samples). The spectral range used for calibration of the flash point was 3944–4769 cm^{-1} and for cetane number was 3500–4678 cm^{-1}. The analytical range for flash point was 24.5–76.5 °C and for cetane number was 37.6–48.9.

The LIBSVM package [29] was employed in this study to develop v-SVR models and the genetic algorithm from Matlab toolbox was applied for parametric optimization. The Matlab 7.8, 64 bits, in a Windows 7 system was used in all calculations.

Different data preprocessing [30,31] were tested in order to chose that which provides the better model development using the PLS and v-SVR algorithms. The tested preprocessing was: baseline correction, baseline correction and mean centering, standard normal variate (SNV), and baseline correction and first derivative. It is of particular interest to consider the order in which preprocessing is applied. This order is completely customizable by the user, but there are some basic rules to follow. In general, it is desirable to perform "row-wise" (sample-based) methods prior to any "column-wise" (variable-based) methods. A row-wise method is one that acts on each sample one at a time (for example, normalization and derivatization). These methods are typically used to remove unwanted variance from individual samples. The effect on any individual sample is independent of the other samples in the data. In contrast, column-wise methods act on variables of the data (for example, centering and scaling). These methods numerically prepare the data for modeling and are thus usually the last methods applied before modeling. These methods often assume that any variance in the data is of interest and use and, as such, is important [30].

To obtain the v-SVR models, different kernel functions, such as: linear, radial basis function (RBF), polynomial and sigmoid were tested. To build the SVR models the data blocks **X** and **Y** of calibration and validation sets, respectively, were previously scaled between [0,1]. Due to its ability to globally locate the optimized solution the genetic algorithm is applied to optimize the SVM parameters C and v. The LIBSVM default value of γ parameter for RBF kernel was used and the parameters C and v were select by GA with parametric optimizing ranges from 0 to 10^4 and 10^{-4} to 1, respectively. For optimization with GA the following parameters were used: number of 30 individuals and a maximum of 15 generations, since it was observed that with these settings the value of the cross validation error was stabilized, not being improved by

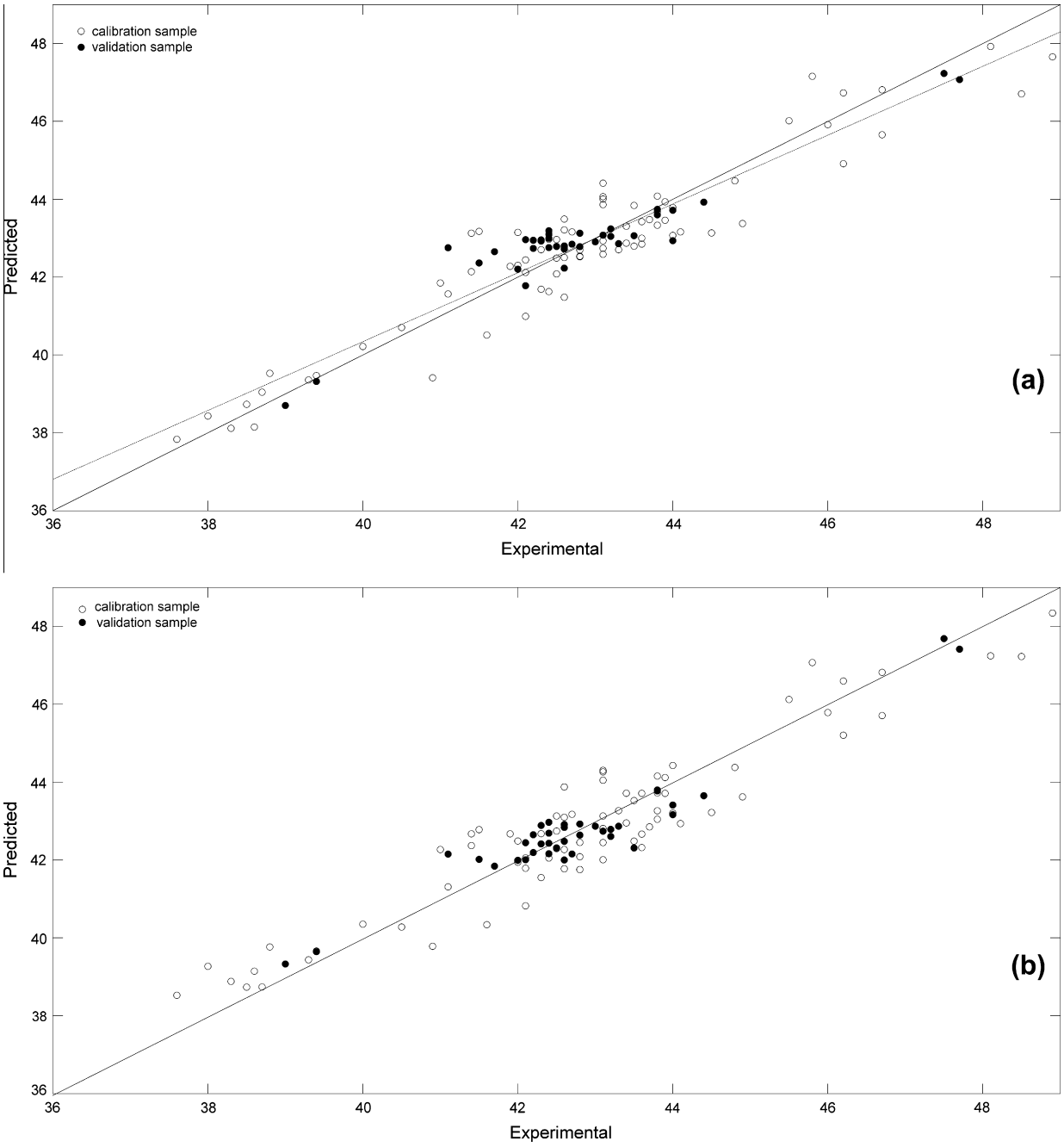

Fig. 6. Cetane number calibration model. Calibration (○) and validation (●) for PLS (a) and SVR (b).

increasing the number of generations. The objective function to be optimized by the GA was the error obtained by cross-validation with five and three subsets of the training set, for flash point and cetane number models, respectively.

As minimizing the error of cross-validation in the training set does not guarantee the optimum condition, a manual grid search was further performed from the values previously selected by the GA.

To evaluate how well the model fits the data was used the root mean square (RMSE), calculated by the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_m - y_p)^2}{n}} \tag{9}$$

where y_p is the predicted value, y_m the measured value and n is the number of samples. The root mean square error of calibration (RMSEC) tells us about the fit of the model to the calibration set. The root mean square error of prediction (RMSEP) was calculated to validation set exactly as in Eq. (9) except that the estimates y_p refer to samples in this data set, which were not involved in either model building or cross-validation.

Table 2
Percentage of validation set samples that fall in the range established by ASTM-E-1655-05 for SVM and PLS models.

	PLS model (%)	SVM model (%)
Flash point	81.2	95.0
Cetane number	100	100

3. Results and discussion

3.1. Flash point

The spectral range between 3944 cm^{-1} and 4769 cm^{-1} was used, which corresponds to the combination bands region. Fig. 2 shows the spectra of 451 samples used in this application. The region used presents several bands assigned to combinations of vibrational modes of the CH bond of methyl and methylene groups and CH bond of aromatic rings [32]. This region allows the flash point calibration because this parameter is related to the proportion of light and heavy fractions of oil in the sample. Long chain molecules of alkanes (paraffins) are major constituents of the lighter fractions and decrease the flash point of diesel oil, while naphthenic and aromatic compounds are mainly present in heavy fractions of petroleum and increase the flash point of diesel oil.

The best result for the PLS model was obtained with SNV as data preprocessing. Three latent variables were used that explain 99.12% of data variance. The results are shown in Table 1, that presents the RMSEC, RMSEP and the R^2 for the linear adjustment between the experimental and predicted values. Fig. 3a illustrates the results of the fitted PLS model.

Although the reproducibility specified by ASTM D56 is 4.3 °C and the repeatability 1.2 °C, the obtained RMSEP value of 3.8 °C for the PLS model developed is not considered satisfactory, since the flash point of diesel oil has a minimum value of 38 °C specified in Brazilian regulations, which means a prediction error of around 10%. Since it is a limiting property for the specification of diesel oil, it is important to obtain a model that provides more accurate results for use by an in-line blending optimizer system. Thus, new prediction models using SVR were proposed, as shown below.

The best result with SVR with the same number of samples used to calibrate and validate the PLS model was obtained using the RBF kernel function and the preprocessed data with baseline correction and mean centering. The parameters C and v were 255.4 and 0.4601, which provides a model that used 239 support vectors. The results are shown in Table 1 and Fig. 3b illustrates the result of SVR model fitted.

The obtained SVR model provides a RMSEP value about 47% better than the RMSEP value obtained with the PLS model. The value of RMSEP of 1.98 °C, makes the SVR model very useful for the purpose of the on-line determination of this quality parameter that is used by the in-line blending optimizer system, since it is below the value specified by flash point reference method reproducibility and must provide accurate results of prediction. The adequate fit of the models can be verified by comparing the RMSEC and RMSEP values obtained for a specific model, which did not differ significantly, indicating that there is no overfitting in the modeling. Also, the number of support vectors used is close to two thirds of the 350 calibration samples, which can be an indication of good model adjustment, because it is considered that the support vectors would used not exceed this fraction of calibration samples [23].

The best fit provided by the use of SVR model, compared to the PLS model can also be verified, through the residual plots shows in Fig. 4 for the calibration sets of PLS and SVR models. Fig. 4a shows a worsening in the predicted values by the PLS model with increasing the flash point, characterizing a nonlinear relationship in the analytical range used. Fig. 4b shows that the SVR model provides a significantly better model in this analytical range.

3.2. Cetane number

The diesel spectra were obtained in the range of 3500–6129 cm^{-1}, which includes the regions of stretch combination bands and the region of first overtone. The absorption bands found in the combination region, mentioned in the previous section, are also useful for determination of the cetane number. Moreover, the absorption bands between 5290 cm^{-1} and 6129 cm^{-1} assigned to the first overtone of stretching modes of the CH bond of methyl and methylene, CH bond of aromatic rings and the CH bond of methyl groups attached to aromatic rings [32], also provide important information for determination of cetane number. Both regions provide important information because the paraffinic compounds increase the values of the cetane number while aromatics decrease these values. The spectra of 114 samples are shown in Fig. 5.

In order to find the best spectral region for the cetane number calibration model several models were tested using: (i) the spectral region mentioned, (ii) only the combination bands (3500–4678 cm^{-1}) and (iii) only the first overtone (5290–6129 cm^{-1}). It was found that both PLS and SVR models gave better results using only the spectral range corresponding to the combination bands region.

The best result with PLS was obtained with baseline correction and mean centered preprocessed data. Five latent variables were used that explain 97.06% of data variance. The results are shown in Table 1 and Fig. 6a illustrates the results of PLS model.

Reference method reproducibility specified by ASTM D613 is 2.8 and the repeatability 0.8, which makes the use of this calibration model possible by taking the comparison with the RMSEP value obtained as a decision parameter. However, it is necessary to consider the importance of this quality parameter in product specification and use for the in-line blending optimizer system. Thus, new prediction models were proposed, aiming to reduce the prediction error, as shown below.

The best SVR model using the same number of samples in calibration and validation sets used for the PLS model was obtained using the RBF kernel function and preprocessed data with baseline correction and mean centering. The parameters C and v were 440.0 and 0.0026, respectively, which provides a model with the use of 11 support vectors. The results are shown in Table 1 and Fig. 6b illustrates the results.

All kernel functions tested, except for the linear kernel function, yielded better results than those obtained with PLS. Here again there is a consistency in the models obtained, which do not use an excessive number of support vectors and give close values of RMSEC and RMSEP, which did not differ significantly, indicating that there is no overfit in the adjusted model. The best SVR model provides a RMSEP value which is about 20% better in comparison to the PLS model value. This value of 0.453, makes the SVR model very useful to be used by the in-line blending optimizer system since it is smaller than the reproducibility specified by the reference method.

3.3. Comparison of NIR results and ASTM specification

In order to verify if the values estimated by the SVR models agree with the specification of the ASTM reference method, a procedure described in ASTM-E-1655-05 was used. In this procedure, one considers the reproducibility, r, of the ASTM reference method and the following equation:

$$y'_i - r < y_i < y'_i + r \tag{10}$$

where y_i is the reference value as obtained by an ASTM reference method and y'_i is the value predicted by the new method. Test results obtained with the same method on identical test items in different laboratories with different operators using different equipment are in reproducibility conditions. The reproducibility value (r) is the difference between two single and independent results, obtained in reproducibility conditions.

Based on the referred ASTM, if 95% or more of the validation set fall in the range determined by Eq. (10) for a given property, then estimates from the model agree with the reference method.

This procedure was applied for both SVR and PLS models and the results are presented in Table 2. The SVR models for flash point and cetane number can be considered to give predicted values that are in the agreement with the ASTM reference method. On the other hand, the PLS model failed in the estimation of flash point.

4. Conclusions

The results show that SVR provides the best regression models in relation to PLS since the SVR can model linear and non linear relationships present in the data sets. On the other hand, it was found that the development of SVR models requires a relatively longer working time, to obtain the parametric optimization, which is fully justified by obtaining more efficient models. The genetic algorithm showed an attractive alternative for SVR optimization, since it can produce appropriate parameters that do not overfit the model.

From the results obtained, the RMSEP values had an improvement of 47% and 21% for the flash point and cetane number, respectively, in relation to the results of PLS models, and all the RMSEP values were smaller than the reproducibility of the corresponding ASTM method. Using ASTM-E-1655-05 to verify if the values estimated by the SVR and PLS models agree with the specification of the ASTM reference method, it was verified that for flash point, only the SVR could be produce predicted values that agree with the ASTM reference method. For cetane number, both SVM and PLS produce results in agreement with the ASTM reference method.

Acknowledgements

The authors would like to thank Petróleo Brasileiro S.A. – PETROBRAS for provide the diesel blends and the NIR spectra for development of this work and CNPq, CAPES and Fapesp for financial support.

References

[1] Chèbre M, Creff Y, Petit N. Feedback control and optimization for the production of commercial fuels by blending. J Process Control 2010;20:441.
[2] Singh A, Forbes JF, Vermeer PJ, Woo SS. Model-based real-time optimization of automotive gasoline blending operations. J Process Control 2000;10:43.
[3] Workman Jr J, Creasy KE, Doherty S, Bond L, Koch M, Ullman A, Veltkamp DJ. Process analytical chemisty. Anal Chem 2001;73:2705.
[4] Shinskey FG. Feedback controllers for the process industries. New York: McGraw-Hill; 1994.
[5] Balabin RM, Safieva RZ. Near-infrared (NIR) spectroscopy for biodiesel analysis: fractional composition, iodine value and cold filter plugging point from one vibrational spectrum. Energy Fuels 2011;25:2373.
[6] Balabin RM, Smirnov SV. Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data. Anal Chim Acta 2011;692:63.
[7] Brudzewski K, Kesik A, Kolodziejczyk K, Zborowska U, Ulaczyk J. Gasoline quality prediction using gas chromatography and FTIR spectroscopy: an artificial intelligence approach. Fuel 2006;85:553.
[8] Balabin RM, Lomakina EI. Support vector machine regression (SVR/LS-SVM) – an alternative to neural networks (ANNs) for analytical chemistry? comparison of nonlinear methods on near infrared (NIR) spectroscopy data. Analyst 2011;136:1703.
[9] Balabin RM, Safieva RZ. Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data. Anal Chim Acta 2011;689:190.
[10] Balabin RM, Safieva RZ, Lomakina EI. Near-infrared (NIR) spectroscopy for motor oil classification: from discriminant analysis to support vector machines. Microchem J 2011;98:121.
[11] Si F, Romero CE, Yao Z, Schuster E, Xu Z, Morey RL, Liebowitz BN. Optimization of coal-fired boiler SCRs based on modified support vector machine models and genetic algorithms. Fuel 2009;88:806.
[12] Balabin RM, Lomakina EI, Safieva RZ. Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. Fuel 2011;90:2007.
[13] Balabin RM, Safieva RZ, Lomakina EI. Gasoline classification using near infrared (NIR) spectroscopy data: comparison of multivariate techniques. Anal Chim Acta 2010;671:27.
[14] Balabin RM, Safieva RZ. Motor oil classification by base stock and viscosity based on near infrared (NIR) spectroscopy data. Fuel 2008;87:2745.
[15] Balabin RM, Safieva RZ. Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. Fuel 2008;87:1096.
[16] Balabin RM, Safieva RZ, Lomakina EI. Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. Chemom Intell Lab Syst 2007;88:183.
[17] Bertran E, Blanco M, Maspoch S, Ortiz MC, Sanchez MS, Sarabia LA. Handling intrinsic non-linearity in near-infrared reflectance spectroscopy. Chemom Intell Lab Syst 1999;49:215.
[18] Geladi P, Kowalski BR. Partial least squares regression: a tutorial. Anal Chim Acta 1986;185:1.
[19] Li H, Liang Y, Xu Q. Support vector machines and its application in chemistry. Chemometr Intell Lab Syst 2009;95:188.
[20] Thissen U, Pepers M, Ustun B, Melssen WJ, Buydens LMC. Comparing support vector machines to PLS for spectral regression applications. Chemom Intel Lab Syst 2004;73:169.
[21] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.
[22] Vapnik V. Statistical learning theory. New York: John Wiley & Sons; 1998.
[23] Ustun B, Melssen WJ, Oudenhuijzen M, Buydens LMC. Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. Anal Chim Acta 2005;544:292.
[24] Smola AJ, Scholkopf B. A tutorial on support vector regression. Stat Comput 2004;14:199.
[25] Chalimourda A, Scholkopf B, Smola AJ. Experimentally optimal v in support vector regression for different noise models and parameter settings. Neural Netw 2004;17:127.
[26] Holland JH. Adaptation in natural and artificial systems. University of Michigan Press; 1975.
[27] Goldberg DE. Genetic algorithms in search, optimization and machine learning. Boston: Addison-Wesley Longman Publishing Co., Inc.; 1989.
[28] Wehrens R, Buydens LMC. Evolutionary optimisation: a tutorial. Trends Anal Chem 1998;17:193.
[29] Chang CC, Lin CJ. LIBSVM: a library for support vector machines; 2001. Software <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
[30] Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W, Koch RS. PLS toolbox version 4.0 for use with Matlab. Wenatchee: Eigenvector research Inc.; 2006.
[31] Romia MB, Bernardez MA. Multivariate calibration for quantitative analysis, in: Infrared spectroscopy for food quality analysis and control. Elsevier Inc.; 2009.
[32] Workman Jr J, Weyer L. Practical guide to interpretive near infrared spectroscopy. Boca Raton: CRC Press; 2008.